Employing Big Data tools to inquire into Albania's residential housing price trends

AN APPLICATION OF PYTHON AND MS EXCEL'S POWER TOOLSET

Introduction

Residential housing is a niche of Albania's real estate market. The latter is one of the most pivotal markets of the domestic economy. The former is a segment focusing on properties reserved to housing only, excluding other real estate such as farm land, forests and pastures, business or commercial space, industrial facilities, etc.

There currently is no official statistics supply on Albania's residential housing market. However, housing property prices for Albania, along with the listing date, pictures, and other relevant information, may be found regularly published online for access to broad viewership.

One such price listing website is njoftime.com. It publishes new listings daily, and it archives its listings up to a 5-year period. The listings published in this website (about 40.000 records dating from January 2017 to February 2022) are used for this research work.

Introduction – research focus

Focus on "Big Data":

- "Big Data" typically refers to the immense volume of Internet-found information, especially UGC (user-generated content).
- Current times' most valuable commodity is INFORMATION. It engulfs us; particularly since the explosion of the World Wide Web, we are surrounded by, what is called, "Big Data";
- The challenge is to be able to:
  a) identify valuable information amidst Big Data clutter,
  b) gather it, and
  c) process it in a systematic way (to then deploy it for traditional use).

Motivation behind this research project:

- Trends of Albania's residential housing market are difficult to track or estimate. As of currently, no official statistics with regard to per-meter-square prices of any real estate properties, are supplied by the Institute of Statistics or other governmental agencies.

Introduction – objective

Objective:

- Estimating a monthly index of the housing properties' per-meter-square price in Albania's capital.
  - *This objective is intended to be achieved by:*
    a) *Making use of website-listed prices for housing property sales in Tirana, as well as Albania's costal areas, and the remainder of its geography;*
    b) *Employing Big Data information extraction and organization tools – i.e: Python and MS Excel's Power toolset – in order to identify, gather, clean, organize, and calculate this website-residing data, so as to synthetize the price index we are interested in.*

Literature review

Housing property prices are typically characterized by wide-ranging variations, which are present because of a multitude of factors that may include:

- building type and material, building amortization, property location,
- floor positioning, climate, building orientation,
- neighborhood placement, nearby facilities, property surface area,
- property architectural arrangement, intent of property use, seller disposition,

…etc.

Furthermore, housing property transactions in Albania are subject to specific taxation, property owners (sellers) are obliged to cover. This property sale tax is calculated considering a reference price that is based on the geographical whereabouts of the property to be sold, and is determined by Albania's Ministry of Finances. The reference price and the tax itself may influence the final property price settlement between the seller and the buyer.

On top of the above, the majority of sellers purchase the services of real estate agencies in order to advertise the sale to as many potential buyers as possible, and these services may also rise even further the final price settlement.

Literature review

Web-scrapped statistical data on residential housing market has increasingly been used in economic research, notably by:

1. Chapelle & Eymeoud (2018), and
2. *Bank of France's* Bricongne, Meunier & Pouget (2021),

…with the goal of acquiring better insights on housing market trends, whether in the house rental or house sale sectors.

This type of data:

a) offers the seller's perspective, through the listings that they, or the real-estate agencies they mandate, post on the Internet (Bricongne et al, 2021);
b) is timely; it publishes property sale prices in real time, thus it pinpoints trends with virtually no delay;
c) beats survey data in terms of financial costs, number of observations, geographical coverage and granularity (Chapelle & Eymeoud, 2018).

Data and methodological framework

The Python-executed data gathering methods used in this study are given in Hajba G. L. (2018)[1] and Kouzis-Loukas D. (2016)[2].

In essence, the procedure requires: [i] within the Python framework, installing, configuring and using a web scraping engine called Scrapy, designed to carry out the electronic data gathering; [ii] identifying the target website, and; [iii] accessing its HTML structure to locate the needed pieces of information intended to be saved for each price listing (observation).

In our case, the 4 pieces of information to be isolated and saved, for each price listing, are:

I. Date (of the listing for sale of the property);
II. Listing agent (he/she who adds the listing for sale);
III. Header (most important details about property, such as listed price and surface square area);
IV. Information (extended details about property, such as location, floor, architectural setup, no. of rooms, etc.).

! Tirane,shes Apartament me hipoteke 2+1+A+BLK Kati 2, 107 m²,160.000 Euro(SECRET GARDEN-TEG)
EuroEstate Albania 11.04.2022, 20:15
Zona/Rruga : TEG
Apartamenti ndodhet ne nje nga rezidencat me te bukura ne Tirane, prane TEG-ut.
Sip totale eshte 107 m2,me ashensor.
Punime super cilesore ,lartesi 3 mt ,veshje me gur,izolim akustik etj
Eshte i organizuar ne :Sallon, 2...

1. *Hajba, Gábor László - Website scrapping with Python, 2018.*
2. *Kouzis-Loukas, Dimitrios - Learning Scrapy: Learn the art of efficient web scraping and crawling with Python, 2016.*

Data and methodological framework

Data are retrieved from njoftime.com, the largest user-generated transactions information website in Albania.

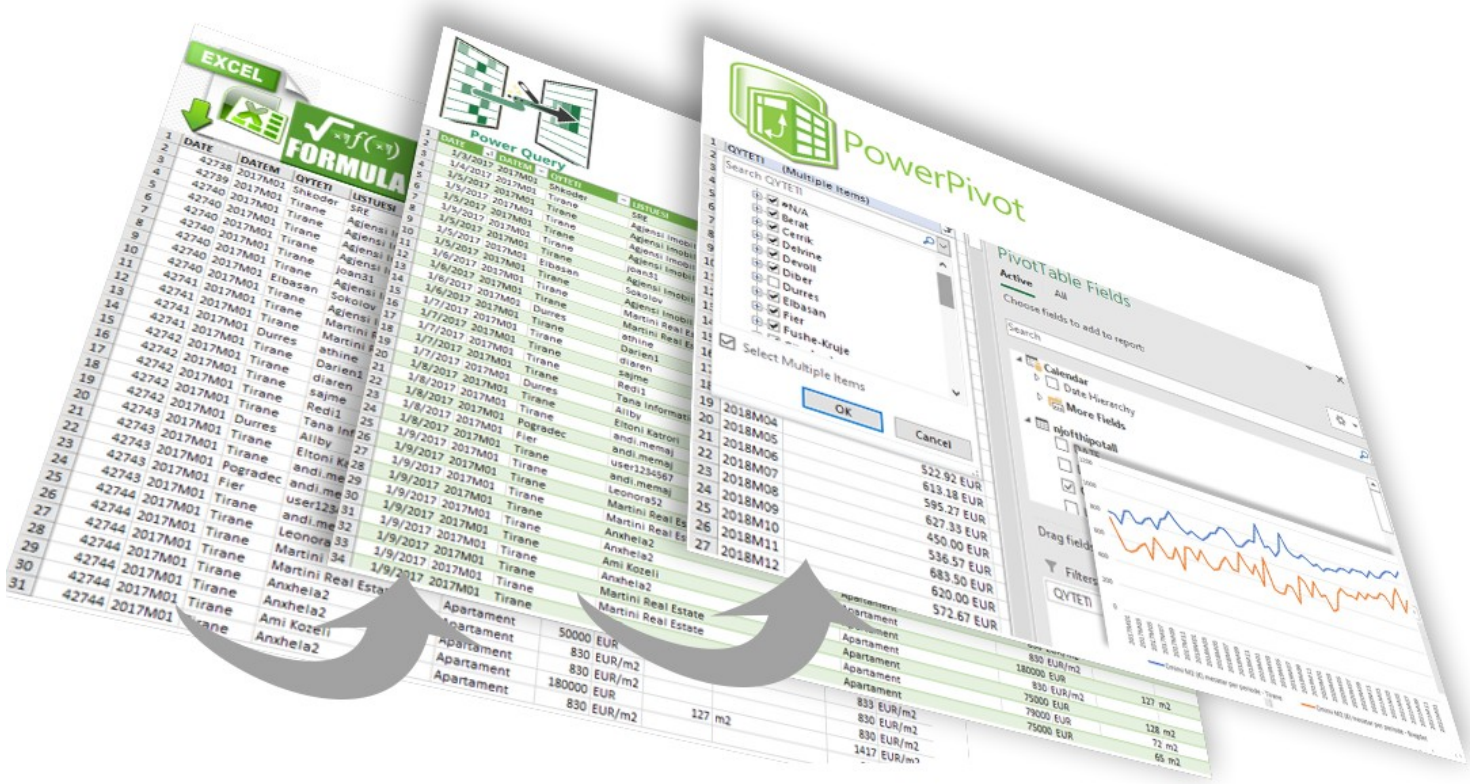
- It yielded 40.491 records of for-sale residence housing listed prices, from January 2017 to February 2022. Of these:
  - *87.7% where prices for sales in Tirana alone;*
  - *8.3% where prices for sale in costal areas;*
  - *3.9% were prices for sales in Albania's inner territory.*
- Python, and its sub-component Srapy, are used to scrape the website content.
- The employed procedure generates a *.json* file, containing all the raw data.

Data and methodological framework

Once the raw, *.json* file—form, data are collected, Excel formulas are applied to filter, clean and organize them in tabular form, with specified variables, no outliers, and no missing values.

Once tabular data sets are ready, through PowerQuery, they are brought together (appended) in one single data framework/model.

This data model is then subjected to PowerPivot and DAX language processing to synthetize the needed results (obtain filtered sets, reports and charts, for the selected time frequency).



Data and methodological framework

Specifics to be clarified regarding the end product (the organized, tabular data):

a) The time frequency is set monthly – but it could've been set up to be quarterly, weekly, or even daily (depending on observations' abundance);

b) There are approximately 648 observations per period (month);

c) The listed residential housing sale data in this study is categorized – referencing the methodology used in the "Real Estate Market" questionnaire, conducted by Bank of Albania's Financial Stability Department – as:

1. Tirana (includes the Tiranë municipality)
2. Albania's costal cities, towns and rural areas (includes the municipalities of):

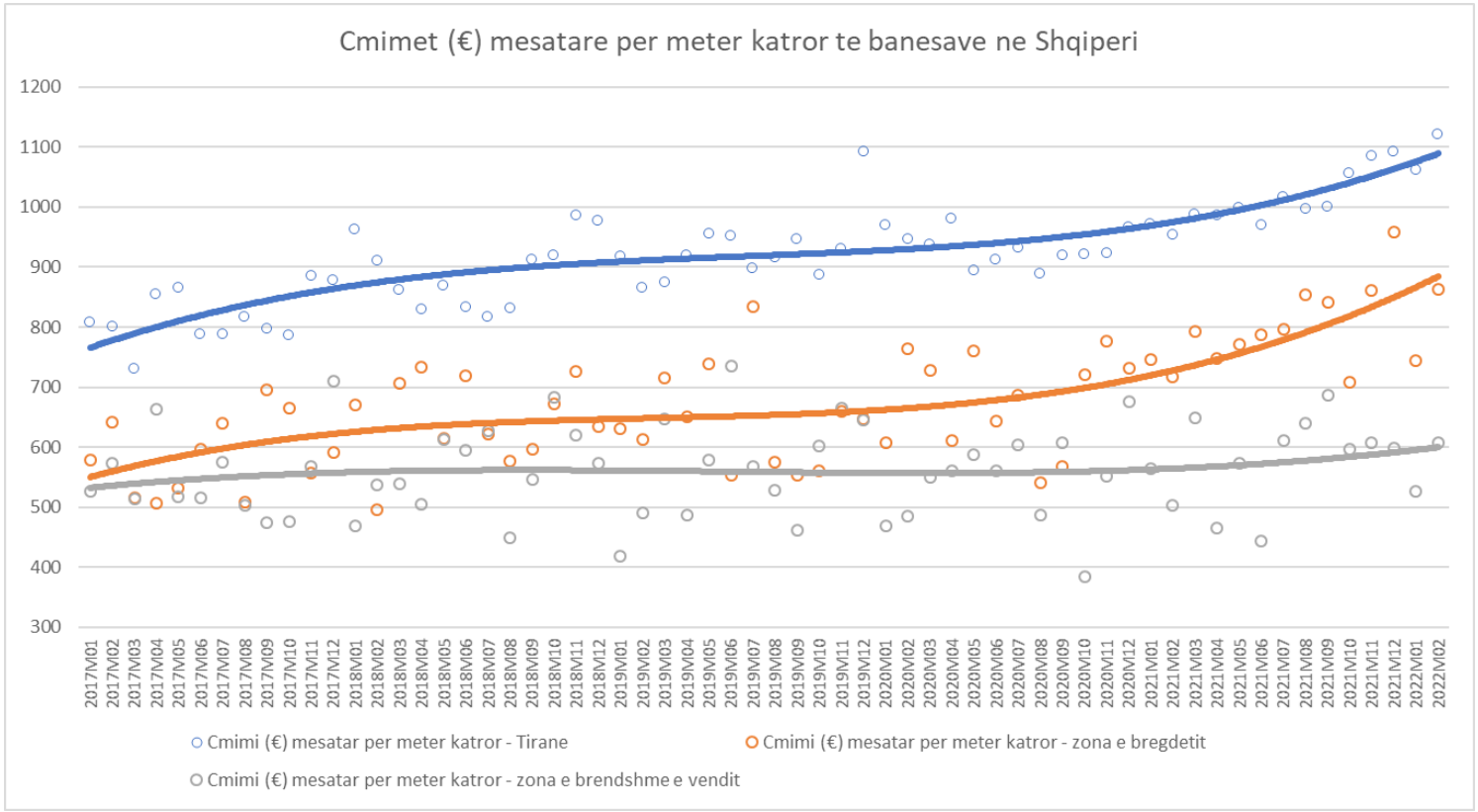
   Durrës, Himarë, Kavajë, Sarandë, Shëngjin, Vlorë.
3. Cities, towns, and rural areas of Albania's inner territory (includes the municipalities of):

   Berat, Cërrik, Delvinë, Devoll, Dibër, Elbasan, Fier, Fushë-Krujë, Gjirokastër, Gramsh, Kamëz, Korçë, Krujë, Kuçovë, Kukës, Laç, Lezhë, Librazhd, Lushnjë, M. e Madhe, Mirditë, Pogradec, Përmet, Shkodër, Tepelenë.

Preliminary results

The graph shows the average, per meter square prices in Euros, for three separate categories of houses locations[3]:

1. those located within the confinements of Tirana Municipality (blue);
2. those located in Albania's costal cities, towns and rural areas (orange);
3. those located in the cities, towns, and rural areas of Albania's inner territory (gray).



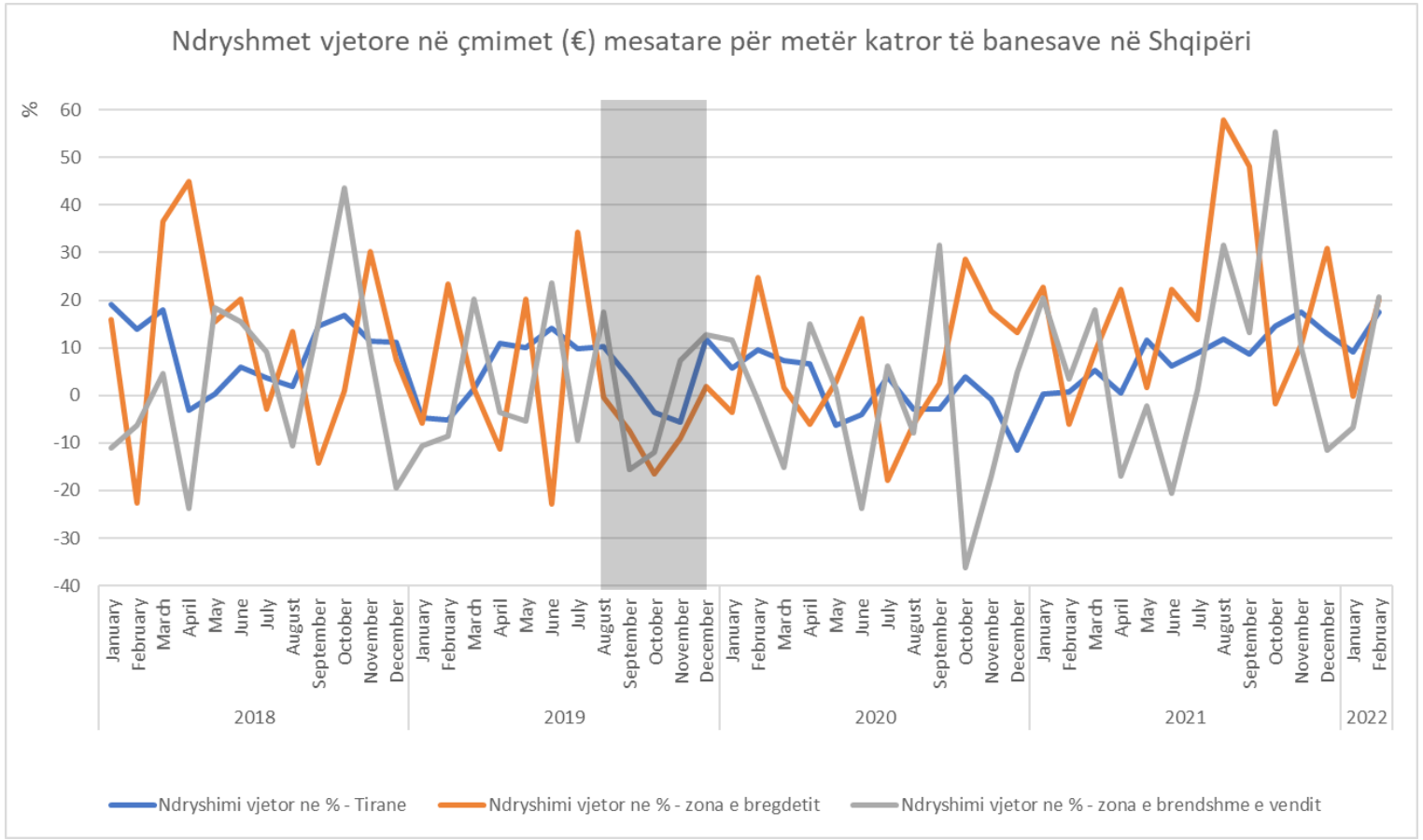
Preliminary results

Price averages, and corresponding YoY percent changes: for year 2021, and latest month data is available (Feb. 2022) are given in the table below:

	2021 price average (€)	2021 YoY percent change in price average	Feb. 2022 price average (€)	Feb. 2020 YoY percent change in price average
Tirana Municipality	1,009.86	8.31%	1,121.65	17.54%
Albania's costal cities, towns and rural areas	793.52	19.01%	862.8	20.25%
Cities, towns, and rural areas of Albania's inner territory	578.23	8.45%	606.71	20.66%

Preliminary results

- YoY percent changes of the average prices display volatility for Albania's costal areas and inner territory.
  - (Seasonality checks yet to be performed).
- All three indicators display negative growth during the second half of 2019 *(the earthquakes natural disaster in Albania)*



Thank you for your attention!

Questions, observations, comments appreciated!

Employing Big Data tools to inquire into Albania's residential housing price trends

AN APPLICATION OF PYTHON AND MS EXCEL'S POWER TOOLSET